



Department of Education, Federal Student Aid

Enterprise Data Dictionary Standards

Version: 1.0

Draft

April 2007



START HERE
GO FURTHER
FEDERAL STUDENT AID™

Table of Contents

Purpose.....	1
Background	2
1.0 Overview	3
1.1 Introduction.....	3
1.2 Benefits of EDD.....	4
1.3 Stakeholders.....	5
1.4 XML Registry and Repository.....	5
1.5 Assumptions.....	5
2.0 EDD Development Standards	7
2.1 Overview.....	7
2.2 Vision.....	7
2.3 EDD Characteristics	8
2.4 EDD Management	9
2.4.1 Management Objectives.....	9
2.4.2 EDD Maintenance.....	9
2.5 Enterprise Data Dictionary - Metadata	10
2.5.1 Basic EDD Metadata (ISO/IEC 11179 Recommendations)	11
2.5.2 Detailed Enterprise Data Dictionary Metadata	12
2.5.3 EDM Enterprise Data Dictionary Metadata.....	14
3.0 EDD Development Guidelines	16
3.1 ISO/IEC 11179 Guidelines	16
3.2 EDM EDD Guidelines	16
4.0 Recommendations	17
4.1 Data Management Tools at Federal Student Aid.....	17
4.2 Additional Recommended Features.....	17
4.3 Open Issues	18
Appendix A. Glossary	19
Appendix B. Abbreviations / Acronyms	20
Appendix C. XML R & R to EDD Mapping	21
Appendix D. Sample EDD (Using ER/Studio).....	24
Appendix E. References.....	26
Appendix F. Recommended EDM EDD Guidelines	27

List of Figures

Figure 1: Data Synchronization data flow.	8
--	---

List of Tables

Table 1: Stakeholders and their needs.	5
Table 2: Basic EDD Metadata.	11
Table 3: Detailed EDD Metadata.....	14
Table 4: EDM EDD Metadata.	15
Table 5: Data Management tools at Federal Student Aid.	17

Document History

Change Number	Date	Reference	A M D ¹	Title or Brief Description	Author	Change Request Number
1.0	04/30/07		M	Final revisions		

¹ A(dd), M(odify), or D(elete)

Purpose

The purpose of this document is to define Enterprise Data Dictionary (EDD) Standards, which will be implemented and used by the Enterprise Data Management (EDM) Team to create and maintain the EDD.

In brief, this document discusses the standards, processes, and procedures needed to create and maintain the EDD at Federal Student Aid. The EDD will be derived from an enterprise-wide metadata repository. All information provided in this document is tool-independent.

Background

Federal Student Aid is engaged in a long-term effort to integrate its processes, data and systems. To better support these business objectives and to emphasize data as an enterprise asset, Federal Student Aid has established a formal Enterprise Data Management (EDM) program. The goal of the EDM program is to consistently define data and make standardized data available across the enterprise by providing information services and data technology expertise to business owners, project managers and architects.

This document is the result of research of best practices and industry standards regarding data dictionaries. It explains what an enterprise data dictionary (EDD) is and what steps are involved in the development. The Enterprise Data Management (EDM) Team commissioned this document as a service for business representatives involved in data dictionary creation and maintenance.

This document outlines the EDD development standards and metadata structure of the dictionary. Comments or suggestions for improvement to these standards are encouraged and should be reported back to the Project Manager for Enterprise Data Management.

1.0 Overview

1.1 Introduction

The EDD is one of the initial components of Enterprise Data Architecture. It is a tool for recording and processing information (metadata) about the data that Federal Student Aid collects and manages.

By definition², a data dictionary is a collection of descriptions of the data objects or items in a data model for the benefit of programmers and others who need to refer to them. A first step in analyzing a system of objects with which users interact is to identify each object and its relationship to other objects. This process is called data modeling and results in a picture of object relationships (entity relationship diagram (ERD)). After each data object or item is given a descriptive name:

- Its relationship is described or it becomes part of a structure that implicitly describes the relationship.
- The type of data (such as text or image or binary value) is described.
- Possible predefined values are listed.
- A brief textual description is provided.

This collection can be organized for reference into a book called a data dictionary. When developing programs that use the data model, a data dictionary can be consulted to understand where a data item fits in the structure, what values it may contain, and, in essence, what the data item means in real-world terms.

It is a catalogue for metadata that can be:

- Used as a central source of information about Federal Student Aid data, a repository for common code.
- Used for data sharing, exchange and integration purposes.
- Referenced during system design, programming, and by actively executing programs.
- Integrated within a database management system (DBMS) or be separate.

The EDD lists the metadata objects, including a complete description of the objects to ensure that they are discrete and clearly understood. Such description must include at least labels (names, titles, etc.) and definitions (or text descriptions), but may also include additional descriptive metadata such as object type, classifications, content data type, rules (business, validation, etc.), and valid and default values. The EDD is the definitive source for the meaning of metadata objects. Some of the benefits of creating an EDD are mentioned in next section.

² From Wikipedia

1.2 Benefits of EDD

An established data dictionary provides numerous benefits for Federal Student Aid:

- **Improved data quality:** Labeling information consistently, with agreed-upon definitions for data elements and a common set of properties for each data element, makes systems and data analysis easier and business intelligence more effective because of access to high data quality information in the EDD.
- **Easy access to trusted data:** Business owners and developers access to validated data including approved definitions and properties supporting Federal Student Aid applications and systems in one central location. As the EDD will be available online, the information provided will always be up-to-date and changes are immediately available to all users. The delay caused by distribution of paper releases is eliminated.
- **Improved documentation and control:** Managing and maintaining all data elements through the EDD ensure consistency and completeness of the data element description.
- **Reduced data redundancy:** Describing data elements and the use of a defined set of properties for each data element reduce or eliminate the creation of redundant data elements. The EDD also allows controlling the addition of new data elements and thereby avoiding duplicates.
- **Reuse of data:** Creating the EDD promotes reuse of data and sharing of information across Federal Student Aid and the community of interest.
- **Consistency in data use:** Implementing a consistent labeling and agreed-upon definition for data elements across applications as well as a defined set of data standards such as naming conventions leads to consistency in data use.
- **Easier data analysis:** Business owners and users might use the EDD as a vehicle for robust query and report generation.
- **Simpler programming:** Using a common set of properties for each data element and consistent labeling of data elements ensure that business and programmer analysts can easily identify relevant data to support implementation of business requirements.
- **Enforcement of standards:** Implementing the EDD with its structure and required data properties establishes an agreed-upon standard that allows for monitoring, controlling, and enforcement of adherence to the standard.
- **Better means of estimating the effect of change:** The EDD will help to identify impact of changes made in the dictionary and its relevant applications and vice-versa.

1.3 Stakeholders

Stakeholders	Needs
Federal Student Aid and Business Owners (System Side)	Improved data quality and consistency in data use
Enterprise Data Management	EDD standards support Enterprise Data Governance and facilitate data exchange
Education Community of Interest	Facilitate data exchange

Table 1: Stakeholders and their needs.

1.4 XML Registry and Repository

Commonly used data elements in various existing applications and common elements identified by business owners are in XML R&R for the Education Community. As described in the document “Federal Student Aid XML Based Approach to Data Management Case Study Final,” the XML R&R framework was created when Federal Student Aid adopted the Service-Oriented Architecture (SOA) with a common data repository for shared data and a metadata repository to orchestrate data management. This framework was developed to address the following strategic drivers for Federal Student Aid:

- Simplify and standardize data exchange with internal and external trading partners.
- Deliver consistent and accurate data across the enterprise-level systems at Federal Student Aid.
- Achieve enterprise-wide efficiencies related to better data-exchange standards and policies.
- Strengthen Federal Student Aid’s relationship with the government and financial aid community data-standards bodies in order to support industry-wide data-exchange standards.

The XML Framework vision was to use XML, via a single set of enterprise and community standards, to simplify and streamline data exchange across postsecondary education. This vision was the foundation of Federal Student Aid’s Enterprise Data Standardization effort under an overall Enterprise Data Management initiative.

The XML Framework enabled Federal Student Aid to realize the benefits of fully integrating XML as an enterprise-wide standard for internal and external data exchange and data storage.

1.5 Assumptions

- The EDD information will be published on the Federal Student Aid intranet and on the PESCS Web site.
- The EDM Team will decide how to manage the EDD as part of EDM-Operations.

- The EDD will initially capture common data elements from XML R&R. Over time, the EDD may expand based on business needs and tool capabilities.
- The EDM Team will collect and provide information about data element usage such as applications that currently use particular data elements. In turn, this information will help with impact analysis, if modifications of a data element are requested.
- The EDM Team will follow the processes and procedures of the governance model for creation and maintenance of the EDD.
- EDM will implement and adhere to Configuration Management for the EDD.

2.0 EDD Development Standards

2.1 Overview

An enterprise-wide data dictionary includes both semantics and representational definitions for data elements. The semantic components focus on creating precise meanings for data elements. Representation definitions include how data elements are stored in a computer structure, such as in integer, string, or date format. Data dictionaries are one step along a pathway of creating precise semantic definitions for an organization.

Initially, data dictionaries are sometimes simply a collection of attributes/database columns and the definitions of the content and data types the attributes/columns contain. Data dictionaries are more precise than glossaries (terms and definitions) because they frequently have one or more representations of how data is structured. Data dictionaries are usually separate from data models because data models often include complex relationships between data elements, which are not captured in the data dictionary. Data dictionaries can evolve into a full ontology when discrete logic has been added to data element definitions.

2.2 Vision

Federal Student Aid's long-term vision is to create an EDD that captures a wide range of information satisfying its business requirements. Section 2.5.2 about EDD metadata lists the recommended EDD data elements. Federal Student Aid's short-term vision is based on the current needs and the availability of the existing tools: XML R & R and ER/Studio's Data Dictionary. Data elements (core components) that are identified and approved by the Postsecondary Electronic Standards Council (PESC) are created in XML R & R under various classifications. This data will be transferred from the XML R & R to the ER/Studio repository as part of the data synchronization effort. Using ER/Studio, Federal Student Aid will develop and maintain an Enterprise Conceptual Data Model (ECDM) and the EDD. The initial version of the EDD will contain information described in the EDM EDD Metadata Template (Section 2.5.3).

Details about the XML R & R and ER/Studio data dictionary mappings are shown in Appendix C and a detailed sample of an EDD data element is shown in Appendix D of this document. A proof of concept to validate these mappings was developed based on the current environment and is referred to as the "Data Synchronization – Case Study".

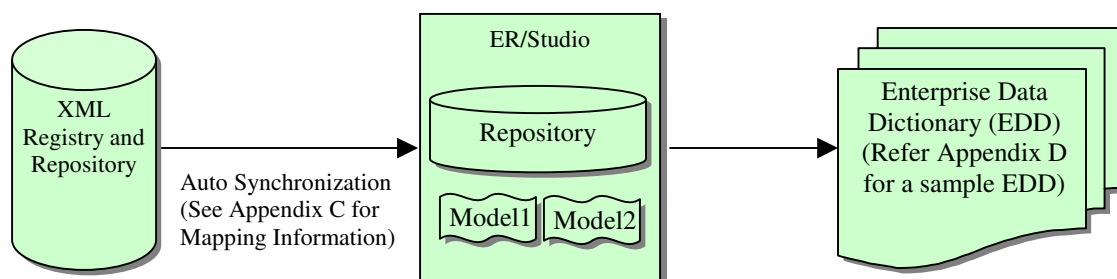


Figure 1: Data Synchronization data flow.

2.3 EDD Characteristics

The following criteria need to be kept in mind when developing an EDD:

- **Consistency:** Corporate data, repositories, etc. are only successful when they resonate with Federal Student Aid and are consistently accessed and maintained within an organization, especially because that data crosses organizational boundaries. An EDD helps to maintain the consistency of corporate data across organizations.
- **Clarity:** An EDD makes data clear and usable for the business user and the developer. It supports efficient and consistent use of the data by both the originators and the various users of the data, regardless of the divisional organization to which they belong. Often, non-standardized data is used because data elements are known within the originating organization without regard to other users outside their organization. The lack of clarity can cause an outside user to misunderstand the meaning, use, or domain of a data element and thus create an erroneous report affecting a management decision.
- **Reusability:** An EDD supports consistent use, which is a key ingredient in the ability of one divisional organization to incorporate work that has already been designed, tested, and approved by the corporation for reuse into its own new development projects. Reinventing the wheel costs money and time. Reusability is enabled by application of standards to produce consistent parts for fitting into future work.
- **Completeness:** An EDD helps analysts know when data is clear, complete, and defined by specifying what completeness means and the steps to develop a complete data structure. Incomplete data properties or descriptions tend to be improperly used and lead to misunderstanding of the data. They can also cost extra time for a developer to make multiple phone calls to clarify and complete the information needed to use the data.

- **Ease of Use for the Developer:** An EDD minimizes development time to have clear and complete definitions/descriptions for the data elements that the programmer must use to create accurately the application functionality.

In addition to being the trusted source for data elements, the EDD shows the relationship between the conceptual view and the data implementation view in the Domain Usage as demonstrated in the sample EDD of Appendix D.

2.4 EDD Management

The management of the EDD falls under the Enterprise Data Management (EDM) Team and will be carried out by the Data Governance and Metadata Manager. With so much detail held in the EDD, it is essential that the data dictionary tool provide an indexing and cross-referencing functionality and possibly a search function. The EDD can produce reports for use by the data administration staff (to investigate the efficiency of use and storage of data), data stewards, management, systems analysts, programmers, and other users. The EDD is one of the resources for data models, databases, and project specific data dictionaries.

2.4.1 Management Objectives

From a management point of view, the EDD should provide facilities for documenting information collected from various applications and models. The EDD should also provide details of usage in various applications, so that analysis and redesign may be facilitated as the environment changes. In general, the EDD will help make information review easier than a paper-based approach by providing cross-referencing and indexing facilities. It also defines standards to be followed in application development.

2.4.2 EDD Maintenance

The EDD evolves based on business needs and technical advances, so it is very important to have a good maintenance plan. This plan needs to identify stewardship, processes for a change in the EDD, and ways to capture all changes made in the EDD under various levels. Part of this maintenance process is the configuration management.

2.4.2.1 Configuration Management

It is essential to keep changes under control, and thus contribute to satisfying quality. Changes need to be auditable and traceable. Configuration management (CM) can be divided into two main areas. The first area of CM concerns the storage of the information produced during the development of EDD: this is sometimes referred to as component repository management. The second area concerns the activities performed during maintenance.

From the perspective of the implementation of a change, the configuration item is the "what" of the change. Altering a specific baseline version of a configuration item creates a new version of the same item. In examining the effect of a change consider both

- (1) What configuration items are affected?, and
- (2) How have the configuration items been affected? A release (itself a versioned entity) may consist of several configuration items.

The set of changes to each configuration item should appear in the release notes describing the differences between the old and the new version of the EDD.

While developing an CM plan for Federal Student Aid's EDD refer to the Institute of Electrical and Electronics Engineering (IEEE) standards and guidelines for additional information:

- IEEE Std. 828-1998 IEEE Standard for Software Configuration Management Plans
- IEEE Std. 1042-1987 IEEE Guide to Software Configuration Management

2.5 Enterprise Data Dictionary - Metadata

The EDD should capture the following information:

- The names associated with each data element (Synonyms)
- A description of each Data Element in natural language (relevant technical information to be added as needed)
- Details of the Registration Authority such as Federal Student Aid, or Postsecondary Electronics Standards Council (PESC)
- Details of the applications/ models that refer to or use each Data Element
- Details about each data element in data processing systems, such as the length of the data element in characters, whether it is numeric, alphanumeric or a different data type, and what logical models include the Data Element
- The validation rules for each Data Element (e.g. permissible values, and range)

Based on the information provided above, three different EDD templates have been designed to satisfy Federal Student Aid requirements at various levels. The **Basic EDD Metadata** template mentioned in Section 2.5.1 is based on ISO/IEC 11179 recommendations (Federal Student Aid follows this standard). Federal Student Aid expects this basic information from an EDD, at a minimum. Section 2.5.2 covers a fully developed **Detailed EDD Metadata** template. This template provides detail information at both the attribute and entity levels. It shows the maximum information an EDD should contain to meet Federal Student Aid current and future requirements. Section 2.5.3 describes the **EDM EDD Metadata** template demonstrating the metadata structure optimized to meet the current Federal Student Aid requirements and is supported by ER/Studio's Data Dictionary tool.

2.5.1 Basic EDD Metadata (ISO/IEC 11179 Recommendations)

The ISO/IEC 11179 standards and guidelines suggest capturing certain data element information to build a data dictionary. Information captured in the following table satisfies the basic ISO/IEC 11179 guidelines. This template can be used as a minimum set of requirements in the creation of EDD.

Data Element	Description
<i>Data Element Name</i>	A unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes.
<i>Identifier</i>	A language independent unique identifier of a data element within a registration authority.
<i>Version</i>	Identification of an issue of a data element specification in a series of evolving data element specifications within a registration authority.
<i>Registration Authority</i>	The organization or body authorized to approve the data elements to include in the Enterprise Data Dictionary.
<i>Synonymous Name</i>	Single word or multi-word designation that differs from the given name, but represents the same data element concept.
<i>Context</i>	A designation or description of the application environment or discipline in which a data item is applied or which it from originates.
<i>Definition</i>	A statement that expresses the essential nature of a data element and its differentiation from all other data elements.
<i>Data Type</i>	A set of distinct values, characterized by properties of those values and by operations on those values.
<i>Maximum Size</i>	Maximum size of data element values
<i>Minimum size</i>	Minimum size of data element values
<i>Permissible Values</i>	The set of representations of permissible instances of the data element, according to the representation form, layout, data type and maximum and minimum size specified in the corresponding attributes. The set can be specified by name, by reference to a source, by enumeration of the representation of the instances or by rules for generating the instances like By list of Values or Range.
<i>Comments</i>	Any additional explanatory remarks on the data element.

Table 2: Basic EDD Metadata.

2.5.2 Detailed Enterprise Data Dictionary Metadata

A Data Dictionary can also provide more complete and detailed information about each data element. This template will give the detail information associated with data element in an EDD. Even though not all the information needs to be captured in an EDD, this template will illustrate how far an EDD can be expanded.

The EDD information can be classified as identification/description, configuration, properties, and association:

- **Identification/ Description:** Contains data element name and definition. This set of fields applies to all data elements (such as definition).
- **Configuration:** Contains essential data element configuration management information provided by the data architect's office. This set of fields applies to all data elements (such as data steward, version, comments, and models).
- **Properties:** Contains attribute and column information (such as data source, data length, value, security, and privacy requirement.)
- **Association:** Contains details of the attributes / columns across the logical and physical data models associated with the data element

Data Element	Description
Identification	
<i>Data Element Name</i>	A unit of data for which the definition, identification, representation, and permissible values are specified by means of a set of attributes.
<i>Synonymous Name</i>	Single word or multi-word designation that differs from the given name, but represents the same data element concept.
<i>Context</i>	A designation or description of the application environment or discipline in which a Data Item is applied or from which it originates.
<i>Definition</i>	A Statement that expresses the essential nature of a data element and its differentiation from all other data elements.
<i>Comments</i>	Any additional explanatory remarks about the data element.
Configuration	
<i>Registration Authority</i>	The organization or body authorized to approve the data elements to include in the Enterprise Data Dictionary.
<i>Identifier</i>	A language-independent unique identifier of a data element within a registration authority.
<i>Version</i>	Identification of an issue of a data element specification in a series of evolving data element specifications within a registration authority.
<i>Current Version</i>	Latest (most recent) version of this data element

Data Element	Description
<i>Phase</i>	Identifies the phase in which the data element is developed or modified.
<i>Status</i>	Possible values: requested, approved, etc –
<i>Status Date</i>	Date the status came into effect
<i>Model Creator</i>	Individual/Organization creating the data model introducing this new data element (author)
<i>Steward</i>	Individual within Federal Student Aid acting as Data Steward
<i>Using Model(s)</i>	Active data model(s) where this data element is part of the data model
<i>Subject Area(s)</i>	Identifies the business context (e.g. Organization. Aid)
<i>Source System</i>	System that determines the need for this data element (e.g. XML R&R requires a new core component that is currently not used in any application, but will be in the future)
<i>Forms</i>	List of forms in which the data element is included, such as data entry forms, report forms, and others
Properties	
<i>Parent Name</i>	Name of the parent entity or association
<i>Parent Identifier</i>	Unique identifier of parent data element
<i>Data Type</i>	A set of distinct values, characterized by properties of those values and by operations on those values.
<i>Maximum Size</i>	Maximum size of data element values
<i>Minimum Size</i>	Minimum size of data element values
<i>Format</i>	How the content of this data element is to be entered/presented to the user (e.g. DD-MM-YYYY for date, or 999-99-9999 for SSN)
<i>Uniqueness</i>	To what extent is an entry in this field unique: <ul style="list-style-type: none"> • Absolute • Unique throughout the entire database • Unique within this table • Unique within the document • N/A – Not Applicable – Not unique
<i>Permissible Values</i>	The set of representations of permissible instances of the data element, according to the representation form, layout, data type, and maximum and minimum size specified in the corresponding attributes. The set can be specified by name, by reference to a source, by enumeration of the representation of the instances or by rules for generating the instances

Data Element	Description
	like By list of Values or Range.
<i>Security and Privacy Requirement</i>	Editable or read-only
<i>Optional</i>	Determines whether a data element is mandatory or optional
<i>Protected (FSA Only)</i>	Data values: yes/no. This flag indicates whether the data element (metadata) and its content are confidential within Federal Student Aid or they can be shared with business partners (e.g. data sharing with community of interest through XML)
Association	
<i>Application Name</i>	Application in which the data element is used.
<i>Entity/Class Name</i>	Any concrete or abstract thing that exists, did exist, or might exist, including associations among these things.
<i>Table Name</i>	Name of the table where the information is captured within the database management system
<i>Attribute Name</i>	Attribute Name for the relevant properties or characteristics of an entity. In the physical model, attributes are represented as table columns.
<i>Column Name</i>	Columns Name for the relevant properties or characteristics of a table.

Table 3: Detailed EDD Metadata.

2.5.3 EDM Enterprise Data Dictionary Metadata

The following template identifies the optimum information that an EDD should capture in order to satisfy EDM business goals and community of interest requirements:

Data Element	Description
<i>Data Element Name</i>	A unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes.
<i>Definition</i>	A statement that expresses the essential nature of a data element and its differentiation from all other data elements.
<i>Context</i>	A description of the application environment or discipline in which a name is applied or from which it originates.
<i>Version</i>	Identification of an issue of a data element specification in a series of evolving data element specifications within a registration authority.
<i>Registration Authority</i>	The organization or body authorized to approve the data elements to include in the Enterprise Data Dictionary.
<i>Data Type</i>	A set of distinct values for representing the data element value.

<i>Maximum Size</i>	Maximum size of data element values
<i>Minimum size</i>	Minimum size of data element values
<i>Comments</i>	Any additional explanatory remarks on the data element.
Permissible Values	
<i>Permissible Values</i>	The set of representations of permissible instances of the data element, according to the representation form, layout, data type and maximum and minimum size specified in the corresponding attributes. The set can be specified by name, by reference to a source, by enumeration of the representation of the instances, or by rules for generating the instances such as By list of Values or Range.
<i>Description</i>	Description of the value
Entity Information	
<i>Application Name</i>	Application in which the entity is used
<i>Entity Name</i>	Name of the entity in the above mentioned application
<i>Start Date</i>	Date the entity is used in the above mentioned application
<i>End Date</i>	Date the entity is no longer used in the above mentioned application
<i>Used</i>	Flag (Active/Inactive): Whether this entity is currently used in the identified application (for example, if the format of the Social Security Number (SSN) changed from alphanumeric to numeric; now the SSN in alphanumeric format is inactive and the SSN in numeric format is active).

Table 4: EDM EDD Metadata.

Information mentioned below will not be published in the EDD but will be maintained internally for audit purposes:

- Created By
- Created Date
- Modified By
- Modified Date
- Approved By
- Approved Date

3.0 EDD Development Guidelines

This section describes the guidelines relevant for creation and maintenance of the Federal Student Aid EDD.

Together with the standards, guidelines will help facilitate the use of agreed-upon EDD data elements, support sharing of the information, permit easy identification of existing data elements, support reuse, and reduce duplication of data elements.

3.1 ISO/IEC 11179 Guidelines

A uniform approach in data dictionary development avoids fragmentation. In an effort to promote and improve international communications between governments, businesses, and scientific communities, ISO and IEC have developed standards for specification and standardization of data elements. The ISO/IEC 11179 standard consists of:

- A framework for the generation and standardization of data elements
- A classification of concepts for the identification of domains
- Basic attributes of data elements
- Rules and guidelines for the formulation of data definitions
- Naming and identification principles for data elements
- Registration of data elements

It is the goal of Federal Student Aid to align with the above-mentioned ISO/IEC Guidelines.

3.2 EDM EDD Guidelines

The following guidelines³ are loosely based on American Health Information Management Association guidelines for Data Dictionary development. Federal Student Aid can use these guidelines as a starting point for developing and maintaining Federal Student Aid – specific standards and guidelines covering the full lifecycle of an EDD:

- Planning
- Development
- Implementation
- Maintenance

For review of the full details of these guidelines go to Appendix F of this document. Federal Student Aid will assess the extent to which these guidelines apply.

³ Guidelines suggested in American Health Information Management Association

4.0 Recommendations

As this document was being written, Federal Student Aid had not decided which tool to use or acquire to develop, publish, and maintain the EDD. The following subsections need to be reviewed and modified (as needed) based on the final tool selection and business requirements. If the existing tools (XML and ER/Studio) will be replaced the standards will remain the same, but the processes may be different.

4.1 Data Management Tools at Federal Student Aid

The following tools are currently in place and used by EDM for data management purposes:

Tool Name	Purpose
<i>XML R & R</i>	Tool used to create and maintain core components. A custom-designed front-end application is used to maintain XML R & R. An Oracle database is used to capture the information.
<i>ER/Studio</i>	Tool used to create and maintain the Enterprise Data Dictionary. The repository is maintained in a Microsoft SQL Server database. It serves as the confirmed source for producing/publishing the EDD until a final tool is selected and implemented, because of its reporting capabilities.
<i>Microsoft Excel, PDF & HTML</i>	Tools that could be used to publish and disseminate EDD.

Table 5: Data Management tools at Federal Student Aid.

4.2 Additional Recommended Features

These are some of the additional features and functionality that are recommended to support the EDD:

- Automatic data synchronization between XML R &R, ECDM and EDD, etc.
- Recognition that several versions of the same program or data structures may exist at the same time. (This point needs to be discussed with the stakeholders)
 - Live and test states of the programs or data.
 - Program and data structures, which may be used at different sites.
 - Data set up under different software or validation routine.
- Provision for an interface with ECDM, Combiner and XML R & R.
- Security features such as password protection, to regulate EDD access.
- Generation of update application programs and programs to produce reports and validation routines.
- Implementing search functionality.

4.3 Open Issues

- Will EDD capture only data elements originated by Federal Student Aid or will it also include core components (XML) introduced through PESC that are not relevant to Federal Student Aid business?
- Is it important to understand and regulate the medium in which the data element is presented (for example: forms on the Web, printed application forms, data entry forms)? This information will affect the level of effort introduced by changes and should be considered when changing a data element used in such a form.
- Identify any specific regulations within ED and/or Federal Student Aid that determine:
 - How to manage sensitive data (e.g., SSN of a person)?
 - Is there a concept of “System of Records”?
 - Does this information affect data sharing policies, data access, and archival strategy, or is it irrelevant because we are looking at Metadata?
- Does the synchronization method need to be defined/refined to ensure that all tags (simple and complex) in XML R & R can be represented accurately in ER/Studio (Hierarchical versus relational data structures)?
- Does the EDD contain both XML and relational data types, or only relational data types? Having both types would allow the user to decide whether or not to review the XML relevant information. If not, the user needs to have access to a translation table to ensure proper mapping/translation of this information.
- Determine the frequency of updates to EDD
- Determine the frequency for publication of EDD

Appendix A. Glossary

The following terms are used in this document or are pertinent to its contents.

Column: A set of data values of the same type collected and stored in the rows of a table.

Database: A set of table spaces and index spaces.

Data Element: A generic term for an entity/class, table, attribute, or column in a conceptual, logical, and physical data model.

Enterprise Conceptual Data Model (ECDM): One of the initial components of Enterprise Data Architecture. The first enterprise level data model developed. The ECDM identifies groupings of data important to Lines of Business, Conceptual Entities, and defines their general relationships. The ECDM provides a picture of the data the enterprise needs to conduct its business. (**Reference:** *U.S. Department of Education Enterprise Data Architecture – Enterprise Data Standards and Guidelines.*)

Enterprise Logical Data Model (ELDM): A component of a maturing Enterprise Data Architecture. The second enterprise level data model developed. It is the result of merging application level data model information into the existing Enterprise Conceptual Data Model (ECDM). The ELDM extends the ECDM level of detail. (**Reference:** *U.S. Department of Education Enterprise Data Architecture – Enterprise Data Standards and Guidelines.*)

eXtensible Markup Language (XML): A meta-markup language for describing data elements that is extensible because it does not have a fixed set of tags and elements.

Schema (XML): A definition, written in eXtensible Markup Language (XML) syntax, of constraints for the content type and data type of XML tags.

Schema (Data): Any diagram or textual description of a structure for representing data. (**Reference:** *FSA-EDM*)

Table: A set of related columns and rows in a relational database.

Table Space: A portion of a database reserved for where a table will go. Table structure is the mapping of tables into table spaces.

Tag (XML): The markup portion of an Extensible Markup Language (XML) element surrounding the character data. The name of the tag reflects the content inside the XML element.

Appendix B. Abbreviations / Acronyms

The following abbreviations and acronyms are used herein or are pertinent to content included herein:

Abbreviation / Acronym	Applicable Term
CDM	Conceptual Data Model
ECDM	Enterprise Conceptual Data Model
ED	Department of Education
EDD	Enterprise Data Dictionary
EDM	Enterprise Data Management
ERD	Entity Relationship Diagram
FEA	Federal Enterprise Architecture
FEAF	Federal Enterprise Architecture Framework
FIPS	Federal Information Processing Standards
IEEE	Institute of Electrical and Electronics Engineers
IEC	International Electro-technical Commission
ISO	International Standards Organization
PESC	Postsecondary Electronic Standards Council
R & R	Registry and Repository
SCM	Software Configuration Management
XML	eXtensible Markup Language
XML R & R	XML Registry and Repository for Education Community

Appendix C. XML R & R to EDD Mapping

Using the existing available resources and information from within the Federal Student Aid to generate the EDD. (The information provided is based on a case study for data synchronization between XML R & R and the Data Dictionary in ER/Studio.)

The tools used are:

1. XML Registry & Repository
2. ER/Studio

Mapping Information:

In order to populate the EDD assumptions were made to generate a Sample EDD. The mapping information developed is shown below.

Mapping Data Types between XML and Data Model Disciplines

In order to generate the Enterprise Data Dictionary, assumptions made to convert the data types between XML and Data Model disciplines **[EDM approval pending]**. The following table shows the mapping between XML data types and data model data types.

XML Data Type	Data Model Data Type
Boolean	?
Date	DATE
Datetime	TIME/DATETIME
Decimal	DECIMAL l s
GDay	CHAR n
GMonth	CHAR n
GYear	CHAR n
GYearMonth	CHAR n
Int	INTEGER
String	VARCHAR n
Token	?
UnsignedInt	INTEGER

1. Every unique XML Business Term becomes a Domain Name in the ER/Studio. **[Assumption: the domain name is the data element name - EDM approval pending]**

Domain Name [Business Term]

Load the Business Terms as Domain Names

XML R & R Name	ER/Studio Name	Data Dictionary Name
Sub Classification Name	Domain Folder	
Business Term	Domain Name	Data Element Name
Definition	Definition	Definition
XML Data Type	Data Type	Data Type
Max Length or Total Digits	Width	Width
Fractional Digits	Scale	Scale

2. Every unique enumeration associated with an XML business term becomes a unique reference value name in the ER/Studio. **[EDM approval pending]**

Reference Values [Enumeration]

1. Load the Reference Values by List
2. Load the Reference Values between range
3. Load the Reference Values not between range

List of Values

XML R & R Name	ER/Studio Name	Data Dictionary Name
Enumeration List Name	Reference Values Name	
	Definition	
List Item Key	Value	
List Item Value	Value Description	

Between range [XML R & R Terminology: With in range]

XML R & R Name	ER/Studio Name	Data Dictionary Name
Business Term	Reference Values Name	
	Definition	

Min Inclusive	Minimum Value	
Max Inclusive	Maximum Value	

Not Between Range [XML R & R Terminology: Out of range]

XML R & R Name	ER/Studio Name	Data Dictionary Name
Business Term	Reference Values Name	
	Definition	
Min Exclusive		
Max Exclusive		

3. Bind the **Reference Value Name [Enumeration]** to corresponding **Domain Name [Business Term]**

4. Load the attributes to the corresponding entities.

5. Bind the Domain Name to the Corresponding Attribute.

6. Generate the Data Dictionary using ER/Studio Tool.

A sample section of the generated EDD is provided as **Appendix D**

Appendix D. Sample EDD (Using ER/Studio)

Domain Detail Reports

Country Code

Domain Name	country code			
Domain Folder	PersonDemographicInformation			
Attribute Name	country code			
Column Name	country code			
Base Datatype	VARCHAR	Width	3	Scale
User Datatype				
Domain Definition				
Domain Note				

REFERENCE VALUE

Reference Value Name	Country Code
Description	Country code short forms
Reference Value Type	By List
Values Not Between	NO

Value	Value Description
IN	INDIA
USA	UNITED STATES OF AMERICA

ATTACHMENTS

Name	Current Value
-------------	----------------------

ATTACHMENTS

Name	Current Value
-------------	----------------------

DOMAIN RESTRICTIONS

Check Constraint
Bound Rule

Declared Default

Bound Default

DOMAIN USAGE

Entity Name	Attribute Name	Model
Person	countrycode	Logical
PersonIdentification	country	Logical

Appendix E. References

The following sources contributed to the content and/or formatting included herein:

- Data Standardization and Procedures document.
- Enterprise Conceptual Data Model – Enterprise Database Dictionary (http://connected1.ed.gov/po/ea/docs/ecdm-edd_overview.doc)
- Federal Student Aid XML Based Approach to Data Management Case Study Final
- Data Synchronization – A Case Study, version 1.0
- http://www-css.fnal.gov/dsg/external/oracle_dcm/9iv2/network.920/a96573/glossary.htm
- http://enterprisestorageforum.Webopedia.com/TERM/D/data_dictionary.html
- http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_030582.hcsp?dDocName=bok1_030582
- <http://www.opengroup.org/architecture/togaf8-doc/arch/chap36.html>
- http://oamWeb.osec.doc.gov/docs/CASD/DOC_CSTARS_Data_Dictionary_Enterprise_Standards_1_1_Final.pdf
- ISO/IEC 11179 of Institute of Electrical and Electronics Engineers (www.ieee.org)
- IEEE Std. 828-1998 IEEE Standard for Software Configuration Management Plans
- IEEE Std. 1042-1987 IEEE Guide to Software Configuration Management

Appendix F. Recommended EDM EDD Guidelines

The following guidelines⁴ are loosely based on American Health Information Management Association guidelines for Data Dictionary development.

▪ Planning Process

- There should be adequate funding and staffing with clearly defined roles and responsibilities for development and implementation of EDD.
- A development plan should be created that clearly identifies the scope, needs, and processes that will be used to develop and maintain the EDD. The plan should be presented to appropriate stakeholders for approval.
- A relevant approving authority/board and its members should be identified to approve any changes in the scope or the processes identified in the planning the process.
- Federal Student Aid should identify the types of media (paper, electronic, spreadsheet, relational database) in which the EDD will be developed, published and maintained. The media choice may depend on the complexity of the enterprise system and the availability of resources.
- There should be provisions to ensure that all licensing agreements are in order.
- An implementation plan should be developed and approved that includes archival strategy on how to manage retired data elements.

▪ Development Process

- Design flexibility and growth capabilities (including room for expansion of field values over time) into the data dictionary so that it will accommodate architecture changes resulting from technical advances or regulatory changes.
- Follow established ISO/ IEC 11179 guidelines and rules for metadata registry (data dictionary) construction to promote interoperability and automated data sharing.
- Develop an enterprise data dictionary that integrates all the data elements used across the enterprise.
- Adopt nationally recognized Federal Information Processing Standards, Geographic codes, Coding Standards and normalize field definitions across data sets to accommodate multiple end user needs. (Federal Information Processing Standards (www.itl.nist.gov/fipspubs), Federal Geographic Data Committee (www.fgdc.gov), United States Postal Service (www.usps.gov), National Spatial Data Infrastructure (www.fgdc.gov/nsdi/nsdi.html), International Organization for Standardization (www.iso.org)).

▪ Implementation Process

- Any variations in the implementation of the enterprise data dictionary that are identified in the planning process need to be documented and approved.

⁴ Guidelines suggested in American Health Information Management Association

- A test plan should be developed to ensure that the system implementation supports the enterprise data dictionary. This should include sampling data inputs and outputs for conformance, validity, and reliability. This process should also verify interoperability of systems.
- Implementation should be done based on the identified media in which the EDD will be developed, published, and maintained.
- An archival strategy on how to manage retired data elements needs to be implemented in accordance with the implementation plan.
- Training of staff based on their use of data elements and their definitions.
- **Maintenance Process**
 - Adequate funding and staffing of the EDM Team with clearly defined roles and responsibilities to ensure ongoing maintenance of the enterprise data dictionary.
 - Develop and implement an approval process following Federal Student Aid's Enterprise Operational Change Management (EOCM) and documentation guidelines for all ongoing EDD maintenance.
 - An EDD tool should/must identify and retain details of versions in two different levels:
 - Data Dictionary level
 - Data Element level

The EDD is dynamic and can be affected by new business lines or changes in national standards.